HSC Research Report

HSC/03/02

Methods for determining the presence of periodic correlation based on the bootstrap methodology

Ewa Broszkiewicz-Suwaj*

* Hugo Steinhaus Center, Institute of Mathematics,
Wrocław University of Technology, Poland

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

Methods for determining the presence of periodic correlation based on the bootstrap methodology

Ewa Broszkiewicz-Suwaj
Institute of Mathematics, Wroclaw University of Technology

Abstract

This paper presents methods for detecting the period of non Gaussian PC processes. A new statistic for testing periodic correlation is proposed. It is based on the bootstrap procedure which is used to estimate confidence intervals of coherence statistic. This method is linked to that of Hurd and Gerr based on Goodman's tests so both methodologies are also compared. It is demonstrated that in some situations the new test appears to be better.

Keywords. Periodic correlation, bootstrap, spectral representation.

1 Introduction

We start with the definition of periodic correlation used in this paper.

Definition 1 *A stochastic sequence* $\{X(n), n \in \mathbb{Z}\} \in L_2(\Omega, F, P)$ *is periodically correlated (PC) with period* T *if :*

a) $\mu_X(n) = E(X(n)) = \mu_X(n+T)$

b) $R_X(m,n) = E[(X(m) - \mu_X(m))\overline{(X(n) - \mu_X(n))}] = R_X(m+T, n+T)$

for every $m, n \in \mathbb{Z}$.

It is known that one of the most important problems in the analysis of periodically correlated sequences is the detection of the period T. The period of the process can be detected from the support of the coherence statistic. The main difficulty is that we don't know the distribution of the sample coherence, so we can't calculate confidence intervals.

In this paper we will present a new method for the detection of the period based on the bootstrap methodology, which enables the approximation of confidence intervals of the sample coherence. We will also compare this method with methods presented by Hurd and Gerr [2].

We shall assume henceforth, without loss of generality, that $\mu(n) \equiv 0$, so the correlation and the covariance are identical.

The methodology presented here is based on the spectral representation of periodically correlated sequences. Gladyshev [1] showed that all PC sequences are harmonizable in the sense of Loéve [4], so the covariance $R_X(m,n)$ has representation

$$R_X(m,n) = \int_0^{2\pi} \int_0^{2\pi} e^{im\omega_1 - in\omega_2} r_Z(d\omega_1, d\omega_2) \tag{1}$$

The support of spectral correlation measure $r_Z(d\omega_1, d\omega_2)$ is then contained in the set of $2 \cdot [T] - 1$ diagonal lines

$$S = \bigcup_{k\in\mathbb{Z}} \{[\omega_1, \omega_2] \in [0, 2\pi) \times [0, 2\pi) : \omega_1 = \omega_2 + 2k\pi/T\}$$

where T is the period of the correlation function (see Hurd and Gerr). So follows that the *spectral density* function is given by

$$f_Z(\omega_1,\omega_2) = \begin{cases} f_Z(\omega_1,\omega_2) & [\omega_1,\omega_2] \in S \\ 0 & [\omega_1,\omega_2] \notin S \end{cases}$$

The main statistic , which we consider is the *coherence statistic*

$$|\gamma(\omega_1,\omega_2)|^2 = \frac{|f_Z(\omega_1,\omega_2)|^2}{f_Z(\omega_1,\omega_1) f_Z(\omega_2,\omega_2)} \ . \tag{2}$$

It takes only real values between 0 and 1 and it is easy to see that the statistic is different from zero only if $[\omega_1, \omega_2] \in S$.

2 Coherent and incoherent statistics

To estimate the coherence statistic we use *sample coherence*.

$$|\gamma(p,q,M)|^2 = \frac{|\sum_{m=0}^{M-1} I_N(\omega_{p+m}) \overline{I_N(\omega_{q+m})}|^2}{\sum_{m=0}^{M-1} |I_N(\omega_{p+m})|^2 \sum_{m=0}^{M-1} |I_N(\omega_{q+m})|^2} \tag{3}$$

where $I_N(\omega)$ is the finite Fourier transform

$$I_N(\omega) = \sum_{n=0}^{N-1} X_n e^{-i\omega n} \tag{4}$$

There are also two tests for determining the presence of periodic correlation proposed by Hurd and Gerr [2] based on the sample coherence.

The *coherent statistic* is defined as $|\gamma(0, d, N)|^2$ where N is the dimension of the vector (length of the sample).

The *incoherent statistic* is given by

$$\delta(d, M) = \frac{1}{L+1} \sum_{p=0}^{L} |\gamma(pM, pM + d, M)|^2 \ , \tag{5}$$

where $L = [\frac{N-1-d}{M}]$.

3 Bootstrap methodology

There are various nonparametric methods to estimate densities and confidence intervals. We only use the *Moving Blocks Bootstrap* (MBB) method ([5], [3]), not because none of the other methods would be appropriate or relevant but because it seems most convenient and it also performs well for PC processes.

Let $\{X_1, \ldots, X_N\}$ be the observed time series. We consider following testing problem:

$$H_0 : |\hat{\gamma}(\omega_p, \omega_q)|^2 = 0 \quad \text{versus} \quad H_A : |\hat{\gamma}(\omega_p, \omega_q)|^2 > 0$$

for $p, q \leq N$ and $p \neq q$. Rejection of H_0 for some (p, q) is equivalent to detecting nonstationarity in time series. For periodically correlated time series in every point $[\omega_1, \omega_2] \in S$ hypothesis H_0 should be rejected. To solve the testing problem we need to know the confidence interval. The methodology for computing confidence interval based on bootstrap runs as follows [5]:

1. we construct B bootstrap replications $\{X^*_{1,b}, \ldots, X^*_{N,b}\}$ using the MBB procedure;

2. we compute replication of coherence statistic $|\gamma^*_i(p, q, M)|^2$ in each point (p, q) B times (for each bootstrap sample);

3. approximate bootstrap confidence interval of level α is given by $[0, \hat{c}_\alpha]$ where $\hat{c}_\alpha$ is the quantile of order $1-\alpha$ of the bootstrap distribution (it is $[\alpha \cdot N]$ largest value of sequence $\{|\gamma_1^*(p,q,M)|^2, \ldots, |\gamma_B^*(p,q,M)|^2\}$).

Knowing the solution of this problem we can easily produce two-dimensional diagonal spectral coherence plots for confidence level α. The main disadvantage of the diagonal spectral coherence plot is interpretation problem. Without having any fit level measure it is difficult to say how strong this plot fits that of spectral mass location for periodically correlated sequences. As measure of fitness we propose the statistic given by

$$MoF(d,M) = \frac{1}{N}\sum_{p=1}^{N} \kappa_\alpha(p, p+d, M) \tag{6}$$

where

$$\kappa_\alpha(p,q,M) = \begin{cases} 1 & |\gamma(p,q,M)|^2 \geq \hat{c}_\alpha, \text{ (when } H_0 \text{ is rejected)} \\ 0 & |\gamma(p,q,M)|^2 < \hat{c}_\alpha \end{cases} \tag{7}$$

We called this statistic the *measure of fitness (MoF)*. It is the function of the difference frequency d and takes only values between 0 and 1 (it can be also interpreted as 0-100%). To compute this statistic we use $I_N(\omega) = I_N(\omega + 2\pi)$.

4 Simulations

The coherence plots presented here were produced using MATLAB from simulated samples of length N=400 and for B=100 bootstrap replications. $N(a,b)$ denotes normal distribution with mean a and standard deviation b, $U(c,d)$ denotes uniform distribution on the interval from c to d.

Peaks in one-dimensional plots (detecting period T) are in points $d_1, 2 \cdot d_1, 3 \cdot d_1, \ldots$ where $d_1 = N/T$ and N is length of sample.

Figure 1. shows that for some processes the MoF statistic is more effective then other one-dimensional statistics. In this case the fast increasing function causes rising of correlation values (even very small). The result is that coherent and incoherent statistics don't detect the real period of the process.

It is important to understand the difference between the two forms of detecting periodic correlation. We easily see that the height of the peak in MoF plots depends on the distinctness of support lines on the two-dimensional plot, so it depends only on the presence of periodic correlation. Statistics

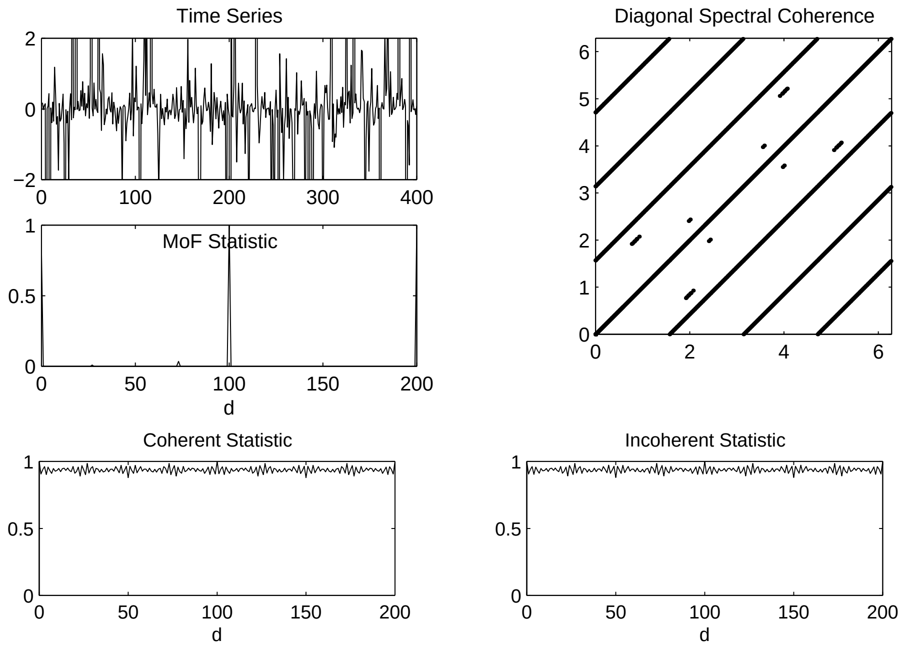

Figure 1: $X_n = U_n \cdot exp(8 \cdot (1 + sin(2\pi n/4) \cdot N_n)$; N_n are $N(0, \frac{4}{5})$ i.i.d., U_n are $U(\frac{1}{2}, \frac{1}{2})$ i.i.d., M=20, $\alpha = 0.01$

presented by Hurd and Gerr depend on the relationship between deterministic and random components (on strength of correlation). This is the main reason why sometimes those statistics are worse.

5 Conclusion

The new method (the MoF statistic) presented here has some disadvantages. Using bootstrap procedure we can only approximate real distribution or confidence interval and there is a possibility for appearance of some errors. Another disadvantage is the computational complexity. However, the examples suggest that it may be more effective than other known methods. Therefore bootstrap methodology appears to be a good tool for detecting the period of periodically correlated processes.

References

[1] Gladyshev E.G. (1961): *Periodically correlated random sequences*, Sov. Math. 2, 383-388.

[2] Harry L. Hurd, Neil L. Gerr (1991): *Graphical Methods for determining the presence of Periodic Correlation*, Journal of Time Series Analysis Vol. 12, No. 4, 337-350.

[3] Leśkow J., Żychaluk K. (2002): *Resampling in assesing stationarity of econometric time series.*

[4] Loéve M. (1991) : *Probability Theory II*, Springer 140-149.

[5] Politis D., Romano J., Wolf M. (1999): *Subsampling* New York, Springer 90-93, 98-99.

HSC Research Report Series 2003

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *Modeling electricity prices: Jump diffusion and regime switching* by Rafał Weron, Michael Bierbrauer and Stefan Trück

02 *Methods for determining the presence of periodic correlation based on the bootstrap methodology* by Ewa Broszkiewicz-Suwaj